ORIGINAL ARTICLE: Clinical Endoscopy

Agreement of site and central readings of ileocolonoscopic scores in Crohn's disease: comparison using data from the EXTEND trial

Paul Rutgeerts, MD, PhD,[1] Walter Reinisch, MD, PhD,[2] Jean-Frederic Colombel, MD,[3] William J. Sandborn, MD,[4] Geert D'Haens, MD, PhD,[5] Joel Petersson, PharmD, PhD,[6] Qian Zhou, PhD,[7] Annalisa Iezzi, MD, PhD,[8] Roopal B. Thakkar, MD[6]

Leuven, Belgium

Background and Aims: Centralized endoscopic scoring may reduce variability, but evidence is lacking in patients with Crohn's disease. We assessed the agreement of endoscopic scorings between site endoscopists and one central reader by using data from the adalimumab Crohn's disease clinical trial EXTEND.

Methods: Agreement between readers for Crohn's Disease Endoscopic Index of Severity (CDEIS)–scored endoscopies from 6 sites and Simple Endoscopic Score for Crohn's Disease (SES-CD)–scored endoscopies from 19 sites in EXTEND was evaluated at baseline and weeks 12 and 52. Agreement on total scores was calculated by using intraclass correlation coefficient (ICC). Kappa statistic or Spearman correlation coefficient measured the agreement between readers for each ileocolonic segment on CDEIS variables including deep ulceration, surface involved, and ulcerated surface and SES-CD variables including ulcerated surface, size of ulcers, and affected surface.

Results: ICCs on mean scores at baseline and weeks 12 and 52 were 0.78, 0.92, and 0.86 (CDEIS), and 0.77, 0.86, and 0.82 (SES-CD), respectively. Site endoscopists consistently reported higher scores. High agreement was observed for most segments and all time points for CDEIS variables and SES-CD large ulcers. Weak agreement occurred for the right side of the colon at all time points for CDEIS deep ulceration and SES-CD large ulcers and at baseline and week 12 for CDEIS ulcerated surface. Fair/moderate agreement occurred for SES-CD ulcerated surface and moderate/high agreement for affected surface for all segments and time points.

Conclusions: Site and central readers showed high agreement on total CDEIS and SES-CD scores overall, whereas variability for individual segments was observed. Weakest agreement occurred at baseline, with a greater difference for SES-CD than for CDEIS score. (Clinical trial registration number: NCT00348283.) (Gastrointest Endosc 2016;83:188-97.)

Crohn's disease (CD) is characterized by chronic inflammation leading to ulceration of the GI mucosa. Current treatment goals for patients with CD include alleviating symptoms and enabling a normal quality of life. Additionally, mucosal healing has been included as an important endpoint in clinical trials assessing the efficacy of anti–tumor necrosis factor therapy in CD.[1-3] Because clinical symptoms and endoscopic activity have not been shown

Abbreviations: CD, Crohn's disease; CDAI, Crohn's Disease Activity Index; CDEIS, Crohn's Disease Endoscopic Index of Severity; EXTEND, Extend the Safety and Efficacy of Adalimumab through Endoscopic Healing; SES-CD, Simple Endoscopic Score for Crohn's Disease; UC, ulcerative colitis.

DISCLOSURES: *P. Rutgeerts has received consultancy fees from AbbVie, Bristol-Myers Squibb, Centocor, Merck, Takeda, and UCB Pharma and speaker fees and research support from AbbVie, Centocor, MSD, and UCB Pharma. W. Reinisch has served as a speaker for Abbott Laboratories, AbbVie, Aesca, Aptalis, Centocor, Celltrion, Danone Austria, Elan, Falk Pharma GmbH, Ferring, Immundiagnostik, Mitsubishi Tanabe Pharma Corporation, MSD, Otsuka, PDL, Pharmacosmos, Schering-Plough, Shire, Takeda, Therakos, Vifor, and Yakult; as a consultant for Abbott Laboratories, AbbVie, Aesca, Amgen, AM Pharma, Astellas, AstraZeneca, Avaxia, Bioclinica, Biogen Idec, Bristol-Myers Squibb, Cellerix, ChemoCentryx, Celgene, Centocor, Celltrion, Covance, Danone Austria, Elan, Falk Pharma GmbH, Ferring, Galapagos, Genentech, Gilead, Grünenthal, ICON, Inova, Janssen, Johnson & Johnson, Kyowa Hakko Kirin Pharma, Lipid Therapeutics, MedImmune, Millennium, Mitsubishi Tanabe Pharma Corporation, MSD, Nestle, Novartis, Ocera, Otsuka, PDL, Pharmacosmos, Pfizer, Procter & Gamble, Prometheus, Robarts Clinical Trials, Schering-Plough, Setpointmedical, Takeda, Therakos,*

(footnotes continued on last page of article)

to correlate well,[4] achieving clinical and endoscopic remission (termed *deep remission*[5]) is evolving as an important target in the treatment of patients with CD.

Two measures of scoring endoscopic disease activity have been partly validated for CD: the Crohn's Disease Endoscopic Index of Severity (CDEIS) and the Simple Endoscopic Score for Crohn's Disease (SES-CD).[6,7] The CDEIS is a scoring system in which 6 endoscopic variables (presence of deep ulcers, superficial ulcers, nonulcerated stenosis, and ulcerated stenosis; proportion of ulcerated surface and of surface involved by disease) are assessed for each of the 5 ileocolonic segments: rectum, sigmoid and left side of the colon, transverse colon, right side of the colon, and ileum.[6] CDEIS scores range from 0 to 44; higher scores indicate more severe disease. The SES-CD is a simpler scoring system based on 4 endoscopic variables (presence and size of ulcers, proportion of surface covered by ulcers, proportion of affected surface, and presence and severity of stenosis), also measured for the 5 ileocolonic segments.[7] Each SES-CD variable is scored from 0 to 3, with the sum of the scores for each variable ranging from 0 to 15 (except for the presence and extent of stenosis, which ranges from 0-11), yielding a total SES-CD score of 0 to 56. As with the CDEIS, higher SES-CD scores indicate more severe disease. CDEIS and SES-CD scores have been shown to be reproducible and well-correlated.[7,8] In contrast, correlation between the Crohn's Disease Activity Index (CDAI) and CDEIS or the CDAI and SES-CD is poor.[6,7]

Although the CDEIS often is considered the standard for evaluating endoscopic disease severity in CD, calculation of the total score is complex, involving multiple measurements, and therefore limits its widespread use. The SES-CD score is easier to calculate and can be an alternative to the CDEIS. Regardless of scoring method, endoscopic assessment of disease activity can be subjective, and interobserver variation may ultimately influence clinical study outcomes. The use of a centralized reviewer in clinical trials has been shown to minimize variability, but the benefit of a centralized reviewer has mostly been observed in ulcerative colitis (UC) clinical trials.[9-11] One study in patients with UC demonstrated greater variability between site and central readers at baseline (screening) than at posttreatment endoscopy. The variability was shown not to be random but rather a systematic upcoding of eligibility scores by site readers compared with central readers, leading to a greater placebo response rate for patients with upcoded scores.[9] Data in patients with CD are lacking. In this post hoc analysis, we determined the agreement between the central and site readers of endoscopies by using data from the CD clinical trial Extend the Safety and Efficacy of Adalimumab through Endoscopic Healing (EXTEND),[1] which demonstrated that adalimumab was more effective than placebo in inducing and maintaining mucosal healing.

It is important to note that a central review committee in EXTEND assessed the primary endpoint of mucosal healing at week 12. If the central reviewer disagreed with a site's assessment, an adjudication process was initiated, involving up to two additional independent blinded reviewers. This post hoc analysis had a different objective, which was to assess the agreement on the components of CDEIS-scored and SES-CD–scored endoscopies between site readers and a single blinded, central reviewer. The site readers were physically on site and performed the readings in real time; the blinded central reviewer in this analysis evaluated all the recorded videos after trial completion.

METHODS

EXTEND trial

Details of the EXTEND trial (NCT00348283) have been published.[1] Briefly, EXTEND was a 52-week, double-blind, placebo-controlled, randomized, maintenance/withdrawal trial to assess the efficacy of adalimumab on mucosal healing in adult patients with moderately to severely active CD, ileocolonic CD for $\geq$4 months, CDAI score of 220 to 450, and mucosal ulceration documented by recorded ileocolonoscopy at screening (performed at study sites). Patients underwent up to 4 endoscopies (screening, week 12, whenever moving to open-label adalimumab after week 12, and week 2).

Ileocolonoscopy scoring

For each ileocolonoscopy, 5 segments (ileum, ascending [right side of the colon], transverse colon, descending [left side of the colon and sigmoid colon], and rectum) were recorded sequentially during endoscope withdrawal. Each segment was recorded for 1 to 2 minutes and was clearly separated. The minimal level of ileocolonic inflammation for patient inclusion in the study was an SES-CD ulcerated surface subscore $\geq$2 in $\geq$1 colon segment (determined by study sites). The primary endpoint of the EXTEND trial was the assessment of mucosal healing, defined as the absence of mucosal ulceration.

For this post hoc analysis, a single central reviewer undertook all SES-CD and CDEIS scoring from videos for comparison with scores from the study sites (Fig. 1). Most study sites (13/19) did not correctly perform and calculate the CDEIS score; however, this was mostly because of incorrect completion of the case report form, which allowed entry of the score as a decimal (correct) or percentage (incorrect).

Patients

Randomized patients who had ileocolonoscopies read by both the blinded central reviewer and the site reader (up to 2 designated endoscopists/site) were included in this analysis. SES-CD scores from all 19 sites and CDEIS scores from the 6 sites that performed and calculated the CDEIS score correctly were used. To assess whether exclusion of 13 sites would be likely to affect outcomes, SES-CD

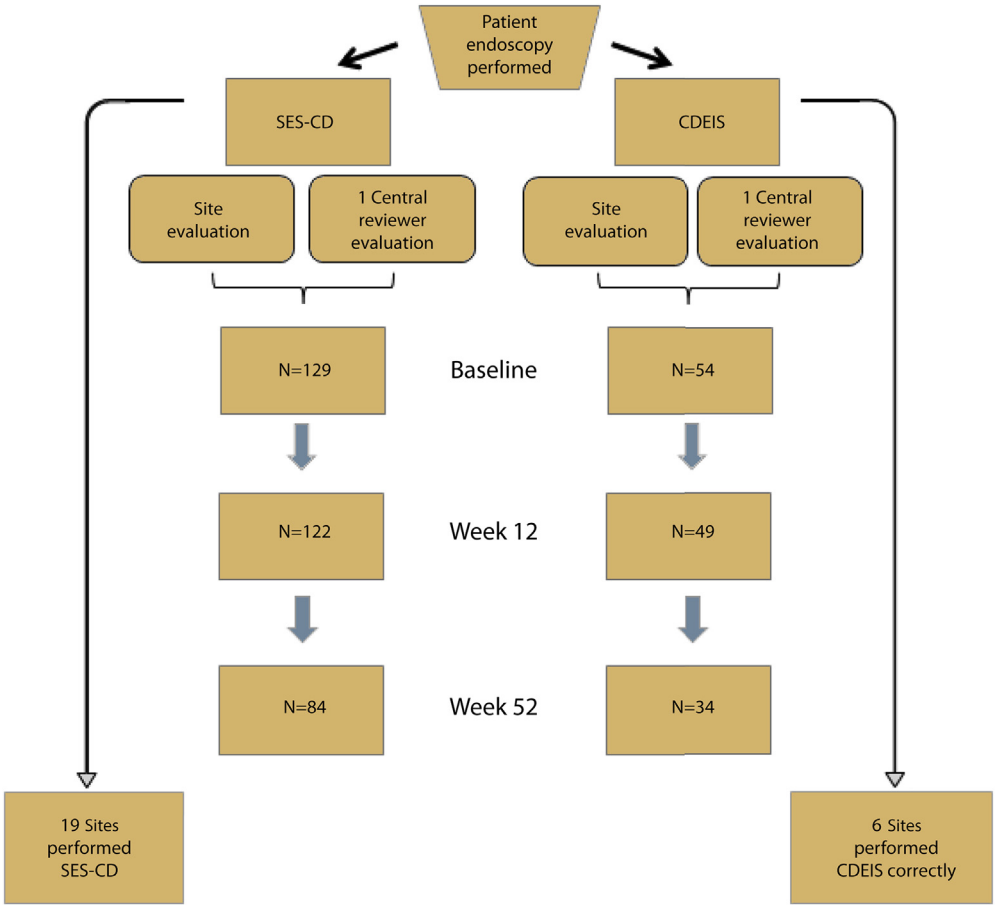

Figure 1. Schematic of the current post hoc analysis assessing agreement between site and central readers in components of SES-CD–scored and CDEIS-scored endoscopies. *SES-CD,* Simple Endoscopic Score for Crohn's Disease; *CDEIS,* Crohn's Disease Endoscopic Index of Severity.

scores also were calculated by using data from the 6 sites that correctly entered the CDEIS values.

Statistical analysis

Demographics were compared between patients with and without CDEIS-scored endoscopies by using the Fisher exact test for sex and race and 1-way analysis of variance for continuous variables. Intraclass correlation coefficients (ICCs) that used a 2-way mixed model were calculated to assess agreement between reviewers, as follows: >0.74 = excellent, 0.60 to 0.74 = good, 0.40 to 0.59 = fair, and <0.40 = poor.[12,13] Interobserver agreement between the central and site reviewers of CDEIS deep ulceration, SES-CD ulcerated surface (subscore ≥2), SES-CD size of ulcers (subscore ≥2), SES-CD affected surface (subscore ≥2), and reduction in total score (CDEIS and SES-CD) from baseline was assessed by using the kappa statistic (κ), as follows: $\kappa = 1$, complete agreement; $\kappa < 1$ to 0.81, almost perfect agreement; $\kappa = 0.8$ to 0.61, high agreement; $\kappa = 0.60$ to 0.21, moderate/fair agreement; $\kappa < 0.2$, low agreement; and $\kappa = 0$, no correlation.[14] The Spearman correlation coefficient (r) was used to measure the agreement between readers of change in total score (CDEIS and SES-CD) from baseline, CDEIS surface involved by the disease, and CDEIS ulcerated surface, as follows: $r = 1$, complete agreement; $r < 1$ to 0.90, very high agreement; $r = 0.89$ to 0.70, high agreement; $r = 0.69$ to 0.30, moderate/fair agreement; $r < 0.29$, low/weak agreement; and $r = 0$, no correlation.[15] When interobserver agreement in each segment was scored, only patients with subscores in all segments at the corresponding time point were included in the analysis.

RESULTS

Patients

Demographics and baseline characteristics of patients randomized in EXTEND have been previously published.[1] In the current analysis, patients were analyzed regardless of randomized treatment; only patients who had endoscopies read by both the central and site reviewers

TABLE 1. Correlation between site and central readers on total CDEIS and SES-CD scores at each visit

	CDEIS			SES-CD			SES-CD (6 sites)		
	Mean (SD) central reading	Mean (SD) site reading	ICC (95% CI)	Mean (SD) central reading	Mean (SD) site reading	ICC (95% CI)	Mean (SD) central reading	Mean (SD) site reading	ICC (95% CI)
Baseline	11.6 (7.4) N = 52	12.9 (6.6) N = 52	0.78 (0.65-0.87)	12.6 (8.3) N = 129	16.8 (8.2) N = 129	0.77 (0.69-0.83)	14.2 (8.4) N = 52	17.4 (8.5) N = 52	0.86 (0.76-0.91)
Week 12	6.9 (6.6) N = 49	7.4 (6.8) N = 49	0.92 (0.86-0.95)	8.0 (7.4) N = 122	9.7 (8.1) N = 122	0.86 (0.81-0.90)	8.0 (7.0) N = 49	9.3 (8.1) N = 49	0.91 (0.85-0.95)
Week 52	4.9 (5.1) N = 34	5.6 (5.8) N = 34	0.86 (0.73-0.93)	6.2 (7.0) N = 84	8.1 (7.5) N = 84	0.82 (0.73-0.88)	6.4 (6.7) N = 34	7.4 (6.8) N = 34	0.83 (0.68-0.91)

CDEIS, Crohn's Disease Endoscopic Index of Severity; *SES-CD,* Simple Endoscopic Score for Crohn's Disease; *SD,* standard deviation; *ICC,* intraclass correlation coefficient based on 2-way mixed model; *CI,* confidence interval.

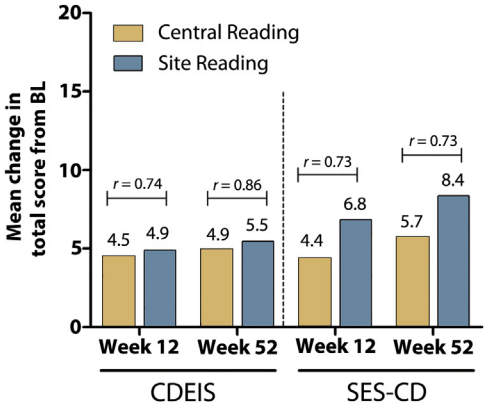

Figure 2. Correlation between central and site readings on mean change from baseline in total CDEIS and SES-CD scores at weeks 12 and 52. Spearman correlation coefficient (*r*) at each time point is shown. *BL,* baseline; *CDEIS,* Crohn's Disease Endoscopic Index of Severity; *SES-CD,* Simple Endoscopic Score for Crohn's Disease.

were included (Fig. 1). Because only 6 sites correctly calculated the CDEIS score, CDEIS-scored endoscopies were available for 52 patients at baseline, 49 at week 12, and 34 at week 52. For SES-CD–scored endoscopies, data from both the central and site reviewers were available for all 129 randomized patients at baseline, 122 at week 12, and 84 at week 52.

Agreement on total scores

Mean CDEIS and SES-CD scores reported by central and site readers and ICC coefficients with 95% confidence intervals (CIs) are shown at baseline, week 12, and week 52 in Table 1. Excellent agreement (ICC >0.74) on total CDEIS and SES-CD scores was observed between central and site readers at all time points, especially after baseline. Site readings were numerically higher than central readings at all time points. When SES-CD data from only the 6 sites that correctly performed and calculated CDEIS values were used to compute SES-CD scores, similar ICC values were obtained as for SES-CD scores from all of the sites.

High agreement was observed between readers when change in total score (CDEIS or SES-CD) from baseline was assessed at weeks 12 and 52 (Fig. 2). The strongest agreement between readers in mean change from baseline in total CDEIS score was observed at week 52 ($r = 0.86$) and at weeks 12 and 52 for total SES-CD score ($r = 0.73$).

The interobserver agreement on the percentage reduction in total CDEIS score from baseline was stronger at week 52 than at week 12 (Table 2). High agreement between central and site readers occurred at week 52 in assessing >25%, >50%, and >75% reduction in total CDEIS score from baseline. At week 12, moderate or fair to high agreement was observed, with the weakest agreement in assessing >25% reduction in total CDEIS score from baseline ($\kappa = 0.54$). For SES-CD scores, agreement between readers was moderate/fair to high at weeks 12 and 52. Agreement was strongest when assessing >75% reduction in SES-CD score ($\kappa = 0.70$) and weakest for >25% reduction in total SES-CD score ($\kappa = 0.53$ at week 12; $\kappa = 0.59$ at week 52).

Agreement in scoring different segments

CDEIS-scored endoscopies. High agreement was observed between the central and site readers in most segments and time points when CDEIS deep ulceration was scored (Table 3). Almost perfect agreement ($0.8 < \kappa < 1$) was observed between readers of the rectum at baseline and week 12, of the sigmoid/left side of the colon at all time points, and in the ileum at baseline and week 52. Perfect agreement ($\kappa = 1$) between readers occurred for the transverse colon at week 12. Moderate and/or fair agreement was observed in the right side of the colon at all time points. Agreement between readers could not be estimated for the rectum at week 52 because the site reader rated all patients as having no deep ulcerations at that time point.

High agreement between readers was observed when CDEIS ulcerated surface in most segments and time points was scored (Table 4). At week 52, agreement was very high ($0.9 \leq r < 1$) in the ileum and complete ($r = 1$) in the

TABLE 2. Degree of agreement (kappa values) between central and site assessments on reduction in total CDEIS and SES-CD scores from baseline by visit

	Week 12			Week 52		
	>25% Reduction	>50% Reduction	>75% Reduction	>25% Reduction	>50% Reduction	>75% Reduction
CDEIS	0.54	0.62	0.64	0.71	0.80	0.75
SES-CD	0.53	0.64	0.70	0.59	0.67	0.70

CDEIS, Crohn's Disease Endoscopic Index of Severity; *SES-CD,* Simple Endoscopic Score for Crohn's Disease.

TABLE 3. Degree of agreement (kappa values) of CDEIS deep ulceration between site and central readings by segment and visit

	Central reading								
	Baseline N = 43			Week 12 N = 38			Week 52 N = 29		
Site reading	No	Yes	κ	No	Yes	κ	No	Yes	κ
Rectum									
No	31	1	0.88	33	1	0.87	27	2	NE
Yes	1	10		0	4		0	0	
Sigmoid and left side of colon									
No	23	1	0.81	31	0	0.91	25	1	0.84
Yes	3	16		1	6		0	3	
Transverse colon									
No	31	3	0.74	34	0	1.0	25	1	0.63
Yes	1	8		0	4		1	2	
Right side of colon									
No	29	2	0.56	31	3	0.37	26	0	0.47
Yes	5	7		2	2		2	1	
Ileum									
No	16	1	0.90	23	4	0.77	19	1	0.92
Yes	1	25		0	11		0	9	

CDEIS, Crohn's Disease Endoscopic Index of Severity; *NE,* not estimable.

transverse colon. Agreement was high ($0.70 \leq r < 0.89$) in the rectum at all time points. Moderate/fair agreement between readers was observed in the sigmoid/left side of the colon at week 52 and in the right side of the colon at weeks 12 and 52. When CDEIS surface involved by the disease was scored, high to very high agreement was observed between readers in most segments and visits (Supplementary Table 1, available online at www.giejournal.org). Complete agreement was observed in the transverse colon at week 52. Moderate/fair agreement between readers was observed in the right side of the colon at week 12 and in the rectum, sigmoid/left side of the colon, and right side of the colon at week 52.

SES-CD–scored endoscopies. Moderate/fair agreement was observed between readers in each segment and at all time points when they scored SES-CD ulcerated surface (subscore ≥2; Supplementary Table 2, available online at www.giejournal.org). The agreement between readers was strongest at week 12 in all segments except the right side of the colon. When SES-CD size of ulcers (subscore ≥2) was scored, a high degree of agreement was observed between readers in most segments and time points, except the right side of the colon, where moderate agreement was observed at all time points (Table 5). Moderate or fair agreement occurred between readers in scoring SES-CD–affected surface (subscore ≥2) in all segments and at all time points. The weakest agreement between readers was observed in the right side of the colon at baseline and in the ileum at baseline and week 12 (Table 6). Agreement on SES-CD presence of narrowing (subscore ≥1) could not be estimated in most segments over time because very few patients had baseline stenosis.

CONCLUSIONS

This analysis evaluated the agreement between a central reader and site endoscopists in scoring components of CDEIS and SES-CD by using data from the EXTEND trial. This was the first study to emphasize central versus site

TABLE 4. Correlation between central and site readings of CDEIS ulcerated surface by segment and visit, Spearman correlation coefficient (*r*)

	Baseline N = 43		Week 12 N = 38		Week 52 N = 29	
	Mean (SD)	*r*	Mean (SD)	*r*	Mean (SD)	*r*
Rectum						
Central	0.67 (1.22)	0.83	0.47 (1.05)	0.85	0.19 (0.36)	0.76
Site	1.26 (2.00)		0.77 (1.90)		0.33 (0.72)	
Sigmoid and left side of colon						
Central	0.88 (1.16)	0.88	0.41 (0.87)	0.85	0.21 (0.60)	0.61
Site	1.75 (2.06)		0.74 (1.73)		0.58 (1.36)	
Transverse colon						
Central	0.52 (0.85)	0.88	0.37 (0.88)	0.85	0.17 (0.46)	1.00
Site	1.33 (2.02)		0.33 (0.84)		0.35 (0.86)	
Right side of colon						
Central	0.43 (0.69)	0.76	0.27 (0.47)	0.51	0.15 (0.46)	0.58
Site	1.30 (1.50)		0.44 (1.01)		0.44 (1.02)	
Ileum						
Central	1.17 (1.15)	0.73	0.69 (0.89)	0.88	0.54 (0.83)	0.91
Site	2.68 (2.67)		1.36 (1.97)		1.45 (1.89)	

CDEIS, Crohn's Disease Endoscopic Index of Severity.

reading correlations for the scoring components of CDEIS and SES-CD by using data from the EXTEND trial. Our results demonstrate that moderate to high agreement occurred between the central and site readers in assessing total endoscopic score and also in evaluating lesions in individual colon segments. Good agreement was observed between readers regardless of scoring method used (CDEIS, ICC 0.78-0.92; SES-CD, ICC 0.77-0.86).[6] Although high agreement between readers in total SES-CD score was observed overall, weaker agreement occurred in scoring SES-CD ulcerated surface in most ileocolonic segments, namely the sigmoid and left side of the colon, right side of the colon, and ileum.

Differences in agreement between the central and site reviewers occurred mostly at study baseline, where site readers reported higher CDEIS and SES-CD scores than did the central reader. This result was not unexpected because the site provided the initial SES-CD assessment. Inflation in scoring by site readers also has been reported for other CD and UC studies.[3,9] It is important to note, however, that the difference in agreement observed at baseline in this analysis is between the site reader and the one central reader who assessed the endoscopy videos after the clinical trial was completed and is not between the site readers and the central review committee that assessed mucosal ulceration during the trial. Over time, the agreement between readers strengthened, but site readers continued to report higher total CDEIS and SES-CD scores than did the central reader at weeks 12 and 52.

One possible explanation for the discrepancy between readers in SES-CD–scored endoscopies may be that the central reviewer was arbitrarily limited to 1-minute recordings of each ileocolonic segment prepared by the site endoscopist. An important point to consider is that the central reviewer determined the extent and degree of ulceration based on the short recorded segments. Aphthous ulcers and smaller extent of ulceration may be more difficult to distinguish in a recorded video than when the endoscopy is performed live and the reader has access to the entire colon in real time. Although the possibility exists that the difference in site and central scoring is related to site readers inflating scores to qualify patients into the trial, the pattern of higher scoring by site readers also was observed at weeks 12 and 52, which argues against upcoding only at baseline for trial eligibility. Rather, these data suggest that determining disease extent and ulceration in real time may be more accurate than assessing it from a recorded video segment. This is especially apparent when the agreement between site and central readers in SES-CD ulcerated surface and in certain segments such as the right side of the colon is assessed, where agreement between readers was quite poor. However, both phenomena may be occurring because the magnitude of difference (ie, the magnitude of upcoding) was greatest at baseline. It should be noted that, in this study, patients were not being re-randomized to maintenance therapy at the primary endpoint (week 12) based on endoscopic response. In contrast, in other study designs in which patients are assessed at the end of an induction trial and then re-randomized to maintenance therapy based on the endoscopic response at the end of induction (yes/no), as have been used in UC, there is a potential incentive for site readers to upcode endoscopic scores so that the patient is judged to not be a success and can then enter an open-label treatment protocol. To be clear, response criteria for

TABLE 5. Degree of agreement (kappa values) of SES-CD large size of ulcers (subscore ≥2) between site and central readings by segment and visit

	Central reading								
	Baseline N = 103			Week 12 N = 95			Week 52 N = 71		
Site reading	No	Yes	κ	No	Yes	κ	No	Yes	κ
Rectum									
No	59	9	0.73	69	9	0.62	55	6	0.72
Yes	4	31		3	14		0	10	
Sigmoid and left side of colon									
No	49	6	0.75	67	8	0.71	53	7	0.65
Yes	7	41		2	18		1	10	
Transverse colon									
No	70	7	0.67	75	5	0.65	63	1	0.93
Yes	6	20		4	11		0	7	
Right side of colon									
No	68	8	0.51	77	8	0.43	62	2	0.46
Yes	11	16		4	6		4	3	
Ileum									
No	54	7	0.74	69	7	0.80	52	3	0.85
Yes	6	36		0	19		1	15	

SES-CD, Simple Endoscopic Score for Crohn's Disease.

endoscopic CD activity have not been validated. Thus, depending on the study design, there is an incentive for site readers to upcode at baseline and possibly at the end of induction. When determining whether to use central readers to qualify patients for a study and/or for determination of the endoscopic endpoint, the risk of site-reader bias (intentional upcoding) must be weighed against the potential advantages of the site reader directly observing the lesions in real time. More data are needed to better understand these two competing phenomena.

Although a high degree of agreement generally was observed between the central and site readers when they assessed different lesions in each segment, assessment of the right side of the colon for CDEIS deep ulceration, CDEIS ulcerated surface, SES-CD size of ulcers, and SES-CD ulcerated surface consistently showed weaker agreement between readers at all 3 time points. Although the right side of the colon may be difficult to view with endoscopy, this cannot completely explain the weak agreement observed between readers because almost perfect agreement was observed in the right side of the colon when CDEIS affected surface (at baseline, week 12, and week 52) was assessed, and high agreement was observed when SES-CD–affected surface (at weeks 12 and 52) was assessed. The reason for the weaker agreement in the right side of the colon is of interest and warrants further analysis.

The results of this study have important implications for clinical trials that include endpoints assessing endoscopic activity. The use of a central reader has been introduced to reduce variability in interpretation of clinical trial results. As stated earlier, this variability could influence the interpretation of treatment efficacy. However, we demonstrated that endoscopies scored by site readers in EXTEND correlated well overall with those scored by the central reader of this study. The correlation between site and central readers on total CDEIS and SES-CD scores observed here is similar to the interrater agreement observed between 4 central readers recently reported by Khanna et al.[16] In the Khanna study, kappa value for interobserver agreement on the CDEIS score was 0.72 and SES-CD score was 0.84,[16] whereas correlation between site and central readers in this analysis ranged from 0.78 to 0.92 and 0.77 to 0.86 for total CDEIS and SES-CD scores, respectively. There was high agreement between readers when >75% reduction in total CDEIS score from baseline was assessed at week 52. Thus, at later time points, our results suggest that central readers and well-trained site readers are similarly effective at scoring endoscopies (understanding the limitation detailed above, that in the current study there was no incentive for the site readers to upcode at the later time points).

The CDEIS score was not reported correctly at 13 of the 19 sites; however, this was mostly because of incorrect completion of the case report form, which allowed entry of the score as a decimal (correct) or a percentage (incorrect). Investigators that use the CDEIS should be made aware of the correct data input format for CDEIS scores, as prospectively defined for the research being conducted.

TABLE 6. Degree of agreement (kappa values) of SES-CD affected surface (subscore ≥2) between site and central readings by segment and visit

	Central reading								
	Baseline N = 103			Week 12 N = 95			Week 52 N = 71		
Site reading	No	Yes	κ	No	Yes	κ	No	Yes	κ
Rectum									
No	62	3	0.65	66	2	0.66	58	3	0.57
Yes	13	25		10	17		4	6	
Sigmoid and left side of colon									
No	52	4	0.60	68	4	0.60	56	1	0.58
Yes	16	31		9	14		7	7	
Transverse colon									
No	65	3	0.55	73	0	0.61	61	0	0.72
Yes	16	19		11	11		4	6	
Right side of colon									
No	72	2	0.49	76	0	0.69	62	0	0.58
Yes	16	13		8	11		5	4	
Ileum									
No	60	6	0.48	77	7	0.49	60	4	0.75
Yes	17	20		4	7		0	7	

SES-CD, Simple Endoscopic Score for Crohn's Disease.

These data also have important implications on the management of patients with active disease. Allez et al[17] demonstrated that the presence of deep ulcers in patients with active ileocolonic CD is a significant short-term and long-term predictor for colectomy. Although a limited number of sites were included in the CDEIS analysis, high agreement between site and central readers on scoring CDEIS deep ulceration was observed at all 3 time points. Thus, our results imply that endoscopists who are experienced in calculating CDEIS can identify patients with more severe disease, who may require more aggressive therapy. For agreement on SES-CD ulcerated surface, it is important to consider that the fair to moderate agreement observed between readers may be influenced by the systematic upgrading of scores by site readers or the limitation of video recordings for the central reviewer, as discussed earlier.

A few limitations of this analysis exist. Fundamentally, there is the inherent risk of bias found with any post hoc analysis. Endoscopies that use CDEIS scoring were included from only the 6 sites that performed CDEIS correctly, whereas SES-CD–scored endoscopies were included from all 19 sites in EXTEND. The small number of sites that reported the CDEIS correctly may limit the ability to compare the interobserver agreement between CDEIS and SES-CD scores. As mentioned earlier, the stronger agreement observed between reviewers who used CDEIS rather than SES-CD scoring could be due to highly experienced and trained site reviewers present at those 6 sites. However, a subanalysis that used only the 6 sites that calculated CDEIS correctly showed similar kappa values to those observed when data from all sites in EXTEND were included in most segments (Supplementary Table 3, available online at www.giejournal.org). Second, although a central review committee with up to 3 central reviewers and up to 2 designated site readers evaluated endoscopy readings in EXTEND, the endoscopy data included in this analysis were rescored by the one blinded central reviewer and were compared with the combined data of the site reviewers. Thus, intraobserver agreement between readers could not be measured here. Last, short video recordings (1 minute) prepared by the site endoscopists could lead to a biased scoring of the endoscopy by the central reader, which may affect the degree of agreement observed between site and central readers.

In conclusion, this study demonstrates that a high degree of agreement occurred overall between site endoscopists and a blinded central reader who used data from the EXTEND trial. However, the evidence of upgrading endoscopic scores at baseline for site endoscopists relative to the central reader resulted in weaker agreement at baseline than at subsequent time points. There was important variability in agreement by anatomic region and for the presence of ulcers, suggesting that certain ileocolonic segments may be more difficult to assess and require expert review.

ACKNOWLEDGMENTS

AbbVie Inc funded the study and analyses and reviewed and approved the manuscript for submission. AbbVie Inc

also participated in the study design, data collection, data management, and data analysis. Medical writing support was provided by Kristina Kligys, PhD, of AbbVie Inc and by Patrick Little, PhD, of Complete Publication Solutions, LLC.

REFERENCES

1. Rutgeerts P, Van Assche G, Sandborn WJ, et al. Adalimumab induces and maintains mucosal healing in patients with Crohn's disease: data from the EXTEND trial. Gastroenterology 2012;142:1102-11 e2.
2. Colombel JF, Sandborn WJ, Reinisch W, et al. Infliximab, azathioprine, or combination therapy for Crohn's disease. N Engl J Med 2010;362:1383-95.
3. Hebuterne X, Lemann M, Bouhnik Y, et al. Endoscopic improvement of mucosal lesions in patients with moderate to severe ileocolonic Crohn's disease following treatment with certolizumab pegol. Gut 2013;62:201-8.
4. Cellier C, Sahmoud T, Froguel E, et al. Correlations between clinical activity, endoscopic severity, and biological parameters in colonic or ileocolonic Crohn's disease. A prospective multicentre study of 121 cases. The Groupe d'Etudes Therapeutiques des Affections Inflammatoires Digestives. Gut 1994;35:231-5.
5. Colombel JF, Rutgeerts PJ, Sandborn WJ, et al. Adalimumab induces deep remission in patients with Crohn's disease. Clin Gastroenterol Hepatol 2014;12:414-22.
6. Mary JY, Modigliani R. Development and validation of an endoscopic index of the severity for Crohn's disease: a prospective multicentre study. Groupe d'Etudes Therapeutiques des Affections Inflammatoires du Tube Digestif (GETAID). Gut 1989;30:983-9.
7. Daperno M, D'Haens G, Van Assche G, et al. Development and validation of a new, simplified endoscopic activity score for Crohn's disease: the SES-CD. Gastrointest Endosc 2004;60:505-12.
8. Sipponen T, Nuutinen H, Turunen U, et al. Endoscopic evaluation of Crohn's disease activity: comparison of the CDEIS and the SES-CD. Inflamm Bowel Dis 2010;16:2131-6.
9. Feagan BG, Sandborn WJ, D'Haens G, et al. The role of centralized reading of endoscopy in a randomized controlled trial of mesalamine for ulcerative colitis. Gastroenterology 2013;145:149-57 e2.
10. Travis SP, Schnell D, Krzeski P, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). Gut 2012;61:535-42.
11. Travis SP, Schnell D, Krzeski P, et al. Reliability and initial validation of the ulcerative colitis endoscopic index of severity. Gastroenterology 2013;145:987-95.
12. Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York: Wiley; 1981.
13. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic 1981;86:127-37.
14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
15. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences, 5th ed. Boston: Houghton Mifflin; 2003.
16. Khanna R, Zou G, D'Haens G, et al. Agreement among central readers in the evaluation of endoscopic disease activity in Crohn's disease. J Crohns Colitis 2014;8:S13-4.
17. Allez M, Lemann M, Bonnet J, et al. Long term outcome of patients with active Crohn's disease exhibiting extensive and deep ulcerations at colonoscopy. Am J Gastroenterol 2002;97:947-53.

Tigenix, UCB, Vifor, Zyngenia, and 4SC; as an advisory board member for Abbott Laboratories, AbbVie, Aesca, Amgen, AM Pharma, Astellas, AstraZeneca, Avaxia, Biogen Idec, Bristol-Myers Squibb, Cellerix, ChemoCentryx, Celgene, Centocor, Celltrion, Danone Austria, Elan, Ferring, Galapagos, Genentech, Grünenthal, Inova, Janssen, Johnson & Johnson, Kyowa Hakko Kirin Pharma, Lipid Therapeutics, MedImmune, Millennium, Mitsubishi Tanabe Pharma Corporation, MSD, Nestle, Novartis, Ocera, Otsuka, PDL, Pharmacosmos, Pfizer, Procter & Gamble, Prometheus, Schering-Plough, Setpointmedical, Takeda, Therakos, Tigenix, UCB, Zyngenia, and 4SC; and has received research funding from Abbott Laboratories, AbbVie, Aesca, Centocor, Falk Pharma GmbH, Immundiagnostik, and MSD. J.-F. Colombel has received consulting and/or lecture fees from AbbVie, ActoGeniX, Albireo Pharma, Amgen, AstraZeneca, Bayer AG, Biogen Idec, Boehringer Ingelheim GmbH, Bristol-Myers Squibb, Cellerix, Centocor, ChemoCentryx, Cosmo Technologies, Danone Research, Elan Pharmaceuticals, Genentech, Giuliani SPA, Given Imaging, GlaxoSmithKline, Hutchison MediPharma, MSD, Neovacs, Ocera Therapeutics, Otsuka America Pharmaceutical, Pfizer, Prometheus Laboratories, Sanofi-Aventis, Schering-Plough, Shire, Synta Pharmaceuticals, Takeda, Teva, Therakos, Tillotts Pharma, UCB Pharma, and Wyeth and has stock ownership in Intestinal Biotech Development, Lille, France. W. Sandborn has received consulting fees from AbbVie, ActoGeniX NV, AGI Therapeutics, Alaven Pharmaceuticals, Alba Therapeutics, Albireo, Alfa Wasserman, Amgen, AM-Pharma BV, Anaphore, Astellas, Athersys, Atlantic Healthcare, Axcan Pharma, BioBalance, Boehringer Ingelheim, Bristol-Myers Squibb, Celgene, Celek Pharmaceuticals, Cellerix SL, Centocor, Cerimon Pharmaceuticals, ChemoCentryx, CoMentis, Cosmo Technologies, Coronado Biosciences, Cytokine Pharmasciences, Eagle Pharmaceuticals, Eisai Medical Research, Elan, EnGene, Eli Lilly, Enteromedics, Exagen Diagnostics, Ferring, Flexion Therapeutics, Funxional Therapeutics, Genzyme, Genentech, Gilead Sciences, Given Imaging, GlaxoSmithKline, Human Genome Sciences, Ironwood Pharmaceuticals, KaloBios Pharmaceuticals, Lexicon Pharmaceuticals, Lycera, Merck Research Laboratories, MerckSerono, Millennium, Nissbin Kyorin Pharmaceuticals, Novo Nordisk, NPS Pharmaceuticals, Optimer Pharmaceuticals, Orexigen Therapeutics, PDL Biopharma, Pfizer, Procter & Gamble, Prometheus Laboratories, ProtAb Limited, Purgenesis Technologies, Receptos, Relypsa Inc, Salient Pharmaceuticals, Salix Pharmaceuticals, Santarus, Schering-Plough, Shire Pharmaceuticals, Sigmoid Pharma, Sirtris Pharmaceuticals, SLA Pharma (UK), Targacept, Teva Pharmaceuticals, Therakos, Tillotts Pharma AG, TxCell SA, UCB Pharma, Viamet Pharmaceuticals, Vascular Biogenics, Warner Chilcott, and Wyeth; lecture fees from AbbVie, Bristol-Myers Squibb, and Janssen and research support from AbbVie, Bristol-Myers Squibb, Genentech, GlaxoSmithKline, Janssen, Millennium, Novartis, Pfizer, Procter & Gamble, Shire, and UCB Pharma. G. D'Haens has received consulting and/or lecture fees from AbbVie, ActoGeniX, AIM, Boehringer Ingelheim GmbH, Centocor, ChemoCentryx, Cosmo Technologies, Elan Pharmaceuticals, Engene, Dr Falk Pharma, Ferring, Galapagos, Giuliani SpA, Given Imaging, GlaxoSmithKline, Janssen Biologics, MSD, Neovacs, Novo Nordisk, Otsuka, PDL Biopharma, Pfizer, Receptos, Salix, Setpoint, Shire Pharmaceuticals, Schering-Plough, Takeda, Tillotts Pharma, UCB Pharma, Versant, and Vifor Pharma; research grants from AbbVie, Janssen, Given Imaging, MSD, Dr Falk Pharma, and Photopill; and speaking honoraria from AbbVie, Tillotts, Tramedico, Ferring, MSD, UCB Pharma, Norgine, and Shire. J. Petersson, Q. Zhou, A. Iezzi, and R. Thakkar are AbbVie employees and shareholders in AbbVie Inc.

0016-5107
http://dx.doi.org/10.1016/j.gie.2015.06.018

Received April 1, 2015. Accepted June 11, 2015.

Current affiliations: Division of Gastroenterology, Department of Internal Medicine, Catholic University of Leuven, Leuven, Belgium (1), Department of Medicine, McMaster University, Hamilton, Ontario, Canada (2), Division of Gastroenterology, Icahn School of Medicine at

Mount Sinai, New York, New York, USA (3), Division of Gastroenterology, University of California San Diego, La Jolla, California, USA (4), Department of Gastroenterology, Academic Medical Center, Amsterdam, The Netherlands and Imelda GI Clinical Research Center, Bonheiden, Belgium (5), Global Medical Affairs Gastroenterology, AbbVie Inc., North Chicago, Illinois, USA (6), Data and Statistical Sciences, AbbVie Inc., North Chicago, Illinois, USA (7), Global Medical Affairs Gastroenterology, AbbVie Inc, Rungis, France (8).

Reprint requests: Paul Rutgeerts, MD, PhD, Department of Gastroenterology, University Hospital Gasthuisberg, Herestraat 49, B-3000, Leuven, Belgium.

SUPPLEMENTARY TABLE 1. Correlation between central and site readings of CDEIS surface involved by the disease by segment and visit, Spearman correlation coefficient (*r*)

	Baseline N = 43		Week 12 N = 38		Week 52 N = 29	
	Mean (SD)	*r*	Mean (SD)	*r*	Mean (SD)	*r*
Rectum						
Central	2.42 (3.56)	0.82	1.47 (3.10)	0.96	0.50 (1.05)	0.66
Site	2.81 (3.52)		1.87 (3.31)		0.66 (1.24)	
Sigmoid and left side of colon						
Central	3.18 (3.69)	0.93	1.45 (2.74)	0.87	0.58 (1.57)	0.55
Site	3.40 (3.55)		1.77 (2.86)		0.78 (1.76)	
Transverse colon						
Central	2.20 (3.57)	0.93	1.35 (3.06)	0.81	0.60 (1.89)	1.00
Site	2.33 (3.47)		1.31 (2.78)		0.53 (1.47)	
Right side of colon						
Central	1.76 (3.07)	0.88	1.33 (2.71)	0.65	0.51 (1.62)	0.45
Site	2.54 (3.38)		1.38 (2.83)		0.62 (1.69)	
Ileum						
Central	3.67 (3.64)	0.91	2.53 (3.46)	0.82	1.92 (2.84)	0.90
Site	4.52 (4.07)		2.51 (3.37)		2.33 (3.29)	

CDEIS, Crohn's Disease Endoscopic Index of Severity.

SUPPLEMENTARY TABLE 2. Degree of agreement (kappa values) of SES-CD ulcerated surface (subscore ≥2) between site and central readings by segment and visit

	Central reading								
	Baseline N = 103			Week 12 N = 95			Week 52 N = 71		
Site reading	No	Yes	κ	No	Yes	κ	No	Yes	κ
Rectum									
No	67	1	0.59	76	1	0.59	59	2	0.53
Yes	16	19		9	9		5	5	
Sigmoid and left side of colon									
No	52	0	0.37	74	1	0.53	59	0	0.45
Yes	32	19		11	9		8	4	
Transverse colon									
No	74	0	0.43	82	0	0.60	63	0	0.37
Yes	19	10		7	6		6	2	
Right side of colon									
No	73	0	0.34	84	0	0.28	67	0	0.39
Yes	22	8		9	2		3	1	
Ileum									
No	51	1	0.28	76	1	0.59	55	1	0.47
Yes	36	15		9	9		9	6	

SES-CD, Simple Endoscopic Score for Crohn's Disease.

SUPPLEMENTARY TABLE 3. Agreement (kappa values) on SES-CD components by segment and visit in all sites and the 6 sites that calculated CDEIS correctly

	Baseline N = 43			Week 12 N = 38			Week 52 N = 29		
	Ulcerated surface	Affected surface	Size of ulcers	Ulcerated surface	Affected surface	Size of ulcers	Ulcerated surface	Affected surface	Size of ulcers
Rectum									
All sites	0.59	0.65	0.73	0.59	0.66	0.62	0.53	0.57	0.72
6 sites	0.64	0.61	0.72	0.84	0.61	0.69	0.65	NE	0.63
Sigmoid and left side of colon									
All sites	0.37	0.60	0.75	0.53	0.60	0.71	0.45	0.58	0.65
6 sites	0.52	0.69	0.86	0.72	0.68	1.0	0.47	0.46	0.71
Transverse colon									
All sites	0.43	0.55	0.67	0.60	0.61	0.65	0.37	0.72	0.93
6 sites	0.42	0.81	0.77	0.84	0.63	0.87	0.47	1.0	1.0
Right side of colon									
All sites	0.34	0.49	0.51	0.28	0.69	0.43	0.39	0.58	0.46
6 sites	0.46	0.78	0.58	NE	1.0	0.47	0.65	1.0	0.36
Ileum									
All sites	0.28	0.48	0.74	0.59	0.49	0.80	0.47	0.75	0.85
6 sites	0.25	0.72	0.80	0.58	0.65	0.83	0.51	0.90	0.93

SES-CD, Simple Endoscopic Score for Crohn's Disease; *CDEIS,* Crohn's Disease Endoscopic Index of Severity; *NE,* not estimable.